Poisson Distributions and MiniBooNE Neutrino Distributions
April 22, 2004
David Finley

Introduction

This memo lays out the conceptual development[i] of a technique for displaying the distribution of n=0,1,2 … neutrinos observed in the MiniBooNE experiment[ii] and comparing it to a prediction based on Poisson distributions. The motivation for this approach is to display as much of the statistical character of the data so that systematic effects can be more easily identified.

The data for this memo are provided in the Analysis Table boodb_anal_pot[iii].

Poisson Statistics

For the purposes of this memo, the Poisson distribution[iv] is given by

$$f(n;nbar) = (nbar)^{n}\, e^{-nbar} / n!\ , \text{ for } n = 0, 1, \ldots$$

where f(n;nbar) is the probability of observing n neutrinos in a span of time during which nbar neutrinos are expected.

A distribution of MiniBooNE neutrinos will converge toward a Poisson distribution if the probability of observing a neutrino in a given period of time is a constant. The usual presentation of a Poisson distribution is done as a function of time, but we do not do that here. Instead, we take advantage of the fact that the data are arranged minute-by-minute in the boodb_anal_pot table. We present the numbers of neutrinos observed in each minute in a histogram, and assert that the distribution will tend to become a Poisson distribution if the number of neutrinos observed in a minute is constant.

The First Idea: Use a Single Expected Value

The first idea for displaying the Poisson statistical characteristics of the data was to use a single value for the expected number of neutrinos per minute for the entire experiment. This is the same as the total number of neutrinos observed divided by the total number of minutes during which the neutrinos were recorded.

On April 6, 2004 the boodb_anal_pot table yielded 228,587 neutrinos distributed in 481,290 minutes. The definition of a "neutrino" in this case are events which pass the “basic neutrino” cuts[v] which require 1) less than 6 veto hits, 2) more than 200 tank hits, and 3) a beam timing cut of split_event >= 4250 nanoseconds and split_event <= 6500 nanoseconds within the 19 microsecond event window.

From these two numbers one can calculate the mean number of neutrinos per minute, and if one then assumes this is the same as the expected number of neutrinos per minute, or nbar, one can calculate a single Poisson distribution. The result is 228,587 / 481,290 =~0.475 = nbar. And one can compare this single Poisson distribution to the observed values as is done in the following table.

n Nus	Observed minutes with n Nus	Poisson Predicted minutes with n Nus	Observed Probability of n Nus	Poisson Predicted Probability of n Nus	Observed Number of Nus for this Row	[Observed minutes - Predicted minutes] / sqrt(Observed-1)
0 Nus	311505	299323	0.6472	0.6219	0	21.8
1 Nu	122709	142162	0.2550	0.2954	122709	-55.5
2 Nus	37256	33760	0.0774	0.0701	74512	18.1
3 Nus	8158	5345	0.0170	0.0111	24474	31.1
4 Nus	1444	635	3.00E-03	1.32E-03	5776	21.3
5 Nus	193	60	4.01E-04	1.25E-04	965	9.6
6 Nus	24	4.77	4.99E-05	9.91E-06	144	4.0
7 Nus	1	0.324	2.08E-06	6.73E-07	7	
>=8 Nus	0	0.019	0	3.99E-08	0	
Sum	481290	481290	1	1	228587	

Note: The first Sum is also the Total Minutes

228587 = Total Observed Number of Neutrinos

0.474946 Total Nus / Total Minutes = nbar

The overall feature of this table is: All of the observed numbers are larger than the predicted values except for the n=1 row. And the discrepancies shown in the final column are significantly larger than the statistical error, sqrt[Observed minutes with n Nus) –1].

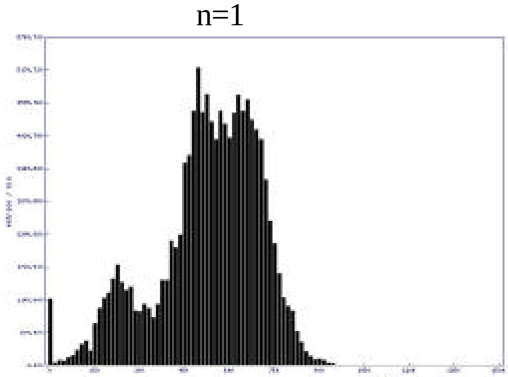
Why the First Idea Is Unsatisfactory and What It Tells Us

The First Idea comparing the data to a single Poisson distribution does clearly show there is an important systematic effect that is not taken into account. The assumption that the probability of producing a neutrino in any given minute is a constant turns out to be too simplistic in the real MiniBooNE world. This assumption is based on many things including assuming the POT/minute is a constant. Although the total number of neutrinos divided by the total POT for the experiment (Nus/POT =~1.1 Nus/E15 POT) is a constant, the POT/minute for each of the minutes not a constant, it is a distribution.

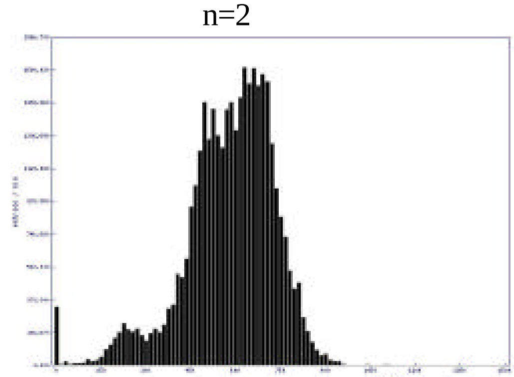
We can imagine the effects of variations in POT/minute when compared to a single Poisson distribution that assumes it is constant, and is taken to be the mean value for POT/minute as in the First Idea. For example, minutes in which the number of POT/minute is less than the mean but still constant would still converge towards a new Poisson distribution, but this new distribution would tend to give more “zero neutrinos” minutes. On the other hand, minutes in which the number of POT/minute is larger than the mean but again constant would also converge to a different Poisson distribution with more “high number of neutrino” minutes. The net result of these other Poisson distributions is to broaden the observed distribution compared to a single Poisson distribution using a single value for nbar. This is what is seen in the table above.

A good way to qualitatively see this is to make a set of histograms showing the distribution of POT/minute for those minutes with n=0,1,2 … neutrinos. The histograms for n=1 and n=2 are shown below, and the distribution for all values of n are given in the Appendix.

n=1



n=2



The horizontal POT/minute scale is the same in both plots: zero to 1500E12. It is apparent that POT/minute is not constant, and varies by more than a factor of two. Also it is interesting to see that the n=2 histogram is skewed to higher POT/minute compared to the n=1 histogram, as one would expect if the number of neutrinos observed in a minute scales with POT/minute. The histograms showing this trend continuing towards higher POT/minute as n grows are given in the Appendix.

A simple correction can be made to compensate for these variations as is described in the next section using the Second Idea.

The Second Idea: Using Minute-by-Minute Expected Values

The Second Idea is to make the assumption that the probability of detecting a neutrino in a particular minute is proportional to the number of protons delivered in that minute, and the constant of proportionality is given by the product of the POT in that minute times the mean value of Nus/POT for the entire experiment. This is a strong assumption, and to the extent that it describes the real data, it could become a convenient tool for sorting out statistical effects from systematic errors.

The first step in implementing the Second Idea is to calculate a Poisson probability distribution $f_i(n; nbar_i)$ for each minute based on $nbar_i$, the expected number of neutrinos for the ith minute. This is given by the product of the POT delivered in the ith minute times Nus/POT for the entire experiment.

The probability for observing n neutrinos in a minute for the entire experiment is the sum over i of the probabilities $f_i(n; nbar_i)$ for each of the minutes. A table of these sums, one sum for each value of n, gives a prediction of the probability for observing n neutrinos in a minute.

Also, since the sum over n of each Poisson distribution for each minute, $f_i(n; nbar_i)$, adds up to unity, the sum of all of the Poisson distributions is equal to the number of minutes. This provides a convenient normalization that can then be used to calculate a prediction of the number of minutes with 0,1,2 … neutrinos in them. And finally, this prediction can be compared to the observed distribution.

The Result of Implementing the Second Idea

The result is given in the following table. This is based the talk Gordon McGregor prepared and showed at the MiniBooNE Collaboration meeting earlier this month.

This table is based on Gordon McGreggor's Talk April 9, 2004 at the MiniBooNE Collaboration Meeting				
n	Observed minutes	Predicted minutes	Neutrinos	Significance
0	310594	310771.9	0	-0.319
1	117226	117051.5	117226	0.510
2	35634	35481.78	71268	0.806
3	7728	7871.05	23184	-1.627
4	1396	1381.82	5584	0.380
5	186	201.59	930	-1.146
6	23	25.27	138	-0.484
7	1	2.79	7	divide by 0
8	0	0.28	0	
9	0	0.02	0	
Sum	472788	472788	218337	
1.12	Nus/E15 POT is used to calculate values for nbar			
The "Significance" is calculated as: (Observed - Predicted) / sqrt(Observed -1)				

The last column shows the number of sigma difference between the observed distribution of minutes with 0,1,2 … neutrinos and the predicted distribution. Sigma is taken to be the square root of the number of Observed minutes. It is quite impressive that the largest number is less than 2 sigma.

For this table, Nus/POT is 218,337/1.94918E20 = 1.120 Nus/E15 POT. The number of neutrinos is lower than in the previous table because two additional "quality" cuts[vi] have been applied. The first quality cut is the "Beam Cut" which requires on a pulse-by-pulse basis[vii]: 1) the two toroids must agree to within 10%, 2) both toroids must be above a threshold of 1E11 protons per pulse, and 3) the horn must be on (indicating a current above 150kAmps). The second quality cut is the "Tank Cut" which is intended to make sure that the time windows in which all the data are collected line up, and to make sure the tank is not latent. (A latency condition occurs when data in the 204.8 microsecond QT circular buffers have been overwritten before retrieval.).

Discussion of the Result of Implementing the Second Idea

First of all, it should be noted that the resulting distribution is not a Poisson distribution if the POT/minute is not constant. Rather it is a sum of Poisson distributions.

Perhaps a tempting overstatement of the assumption is: The probability of seeing a neutrino in a given minute ONLY depends on how many protons were delivered in that minute, AND everything else stays the same. Stated this way, it begs questions like: Does the targeting of the protons vary significantly over the course of the experiment? Do the gains in the phototubes vary significantly over the duration of the experiment? If they do, the assumption of always using the same mean for Nus/POT would no longer be good. (Of course, the ongoing "horn off" data show a large difference in the probability compared to "horn on", and this is an example of a systematic change clearly invalidating the assumption.)

Systematic variations that would escape detection by this analysis are those for which the ratio Nus/POT remains constant. A simple example would be if the toroids begin to read 5% lower AND at the same time the phototube efficiencies also drop by 5%. This particular example is unlikely (and even if it did occur it might not impact certain physics analyses), but other examples may not be so unlikely.

If the comparison produced from this assumption with the data look good, it may also mean that there were no significant systematic variations for the duration of the experiment in the probability of observing a neutrino, other than the number of protons delivered in a given minute. This would be stunning, but possible. However, it would be prudent to apply this analysis to subsets of the data to see if one can perhaps uncover systematic variations that are swamped by looking at the entire data set. For example, a group of five subsets each covering 20% of the minutes would be a good start.

A Final Diversion: Predictions for Finer Grained Time Slices

The boodb_anal_pot table collects the data in to minute long time slices. Because of this, information on the pulse time scale, and on the bunch time scale is not available at this time from the Analysis Tables.

The variations in POT/minute are significant and need to be compensated, as shown above. The largest contribution to this variation comes from the variation in average repetition rate that is used for MiniBooNE. This can be inferred from the weekly average of the "Horn Rate" plot in the MiniBooNE weekly performance plots[viii]. The rate varies from ~1 Hz to 3 Hz. Of course, the POT/minute also depends on variations in "POT per Horn Pulse", but this variation is smaller, from ~3E12 to ~4.5E12. Finally, there are variations in POT/minute due to "Beam Uptime Fraction" and these vary from ~80% to 98%.

Thus, it might be instructive to predict what should be seen on a pulse-by-pulse or even a bunch-to-bunch basis since the "First Idea" should work better than it does on the minute-by-minute basis. Of course one can still apply the Second Idea to the pulse-by-pulse data to remove variations in POT/pulse.

The following table on the left shows the single Poisson distribution one obtains if one assumes POT/pulse = 3E12, and the table on the right shows the case for 4.5E12 POT/pulse.

For	**3.00E+12**	**protons/pulse**	
	and		
	2.50E+20	**POT total**	
	1.1E-15	**Nus/POT**	
One obtains			
3.30E-03	nbar		
8.33E+07	pulses		
2.75E+05	neutrinos		
n	**Predicted Probability**	**Predicted Pulses**	**Neutrinos**
0	0.99670544	83,058,787	0
1	3.29E-03	274,094	274,094
2	5.43E-06	452	905
3	5.97E-09	0.50	1.49
4	4.93E-12	4.10E-04	1.64E-03
Sum	1	83,333,333	275,000

For	**4.50E+12**	**protons/pulse**	
	and		
	2.50E+20	**POT total**	
	1.1E-15	**Nus/POT**	
One obtains			
4.95E-03	nbar		
5.56E+07	pulses		
2.75E+05	neutrinos		
n	**Predicted Probability**	**Predicted Pulses**	**Neutrinos**
0	0.9950622	55,281,235	0
1	4.93E-03	273,642	273,642
2	1.22E-05	677	1,355
3	2.01E-08	1.12	3.35
4	2.49E-11	1.38E-03	5.53E-03
Sum	1	55,555,556	275,000

One clear prediction from these tables is: For the MiniBooNE data sample taken to date, there should be several hundred (452 to 677) pulses with n=2 neutrinos in them, but one pulse or less with n=3 neutrinos or more in them. If this can be checked in the data, meaning if one can look for multiple neutrino interactions within the same beam pulse, it would be an amusing but clear check of the Poisson statistical characteristic of the MiniBooNE data.

As a final exercise, we carry the calculation to the final level, that of a bunch-by-bunch prediction. There are places for each of exactly 84 bunches in the Booster, but the requirements of lowering losses at extraction results in only ~80 of these places actually contiguously occupying beam. Thus, MiniBooNE sees ~80 bunches per pulse; these bunches are separated by ~19 nanoseconds. (In order to check the uniformity of protons/bunch over the entire experiment, one needs to use the resistive wall monitor on a pulse-by-pulse basis.) The results are given in the following two tables.

For	**3.00E+12**	**protons/pulse**	
	and		
	80	**bunches/pulse**	
	2.50E+20	**POT total**	
	1.1E-15	**Nus/POT**	
	One obtains		
	4.13E-05	nbar	
	6.67E+09	bunches	
	2.75E+05	neutrinos	
n	**Predicted Probability**	**Predicted Bunches**	**Neutrinos**
0	0.9999588	6,666,391,672	0
1	4.12E-05	274,989	274,989
2	8.51E-10	5.67	11
3	1.17E-14	7.80E-05	0.00
4	1.21E-19	8.04E-10	3.22E-09
Sum	1	6,666,666,667	275,000

For	**4.50E+12**	**protons/pulse**	
	and		
	80	**bunches/pulse**	
	2.50E+20	**POT total**	
	1.1E-15	**Nus/POT**	
	One obtains		
	6.19E-05	nbar	
	4.44E+09	bunches	
	2.75E+05	neutrinos	
n	**Predicted Probability**	**Predicted Bunches**	**Neutrinos**
0	0.99993813	4,444,169,453	0
1	6.19E-05	274,983	274,983
2	1.91E-09	8.51	17
3	3.95E-14	1.75E-04	0.00
4	6.11E-19	2.71E-09	1.09E-08
Sum	1	4,444,444,444	275,000

These tables indicate there should be several (5 to 9) bunches already in the MiniBooNE data which have produced n=2 neutrinos within the same bunch. Since a bunch is less than a few nanoseconds long, the two neutrinos will be "in time" to better than a few nanoseconds, but they will be distributed transversely over the detector.

Conclusion

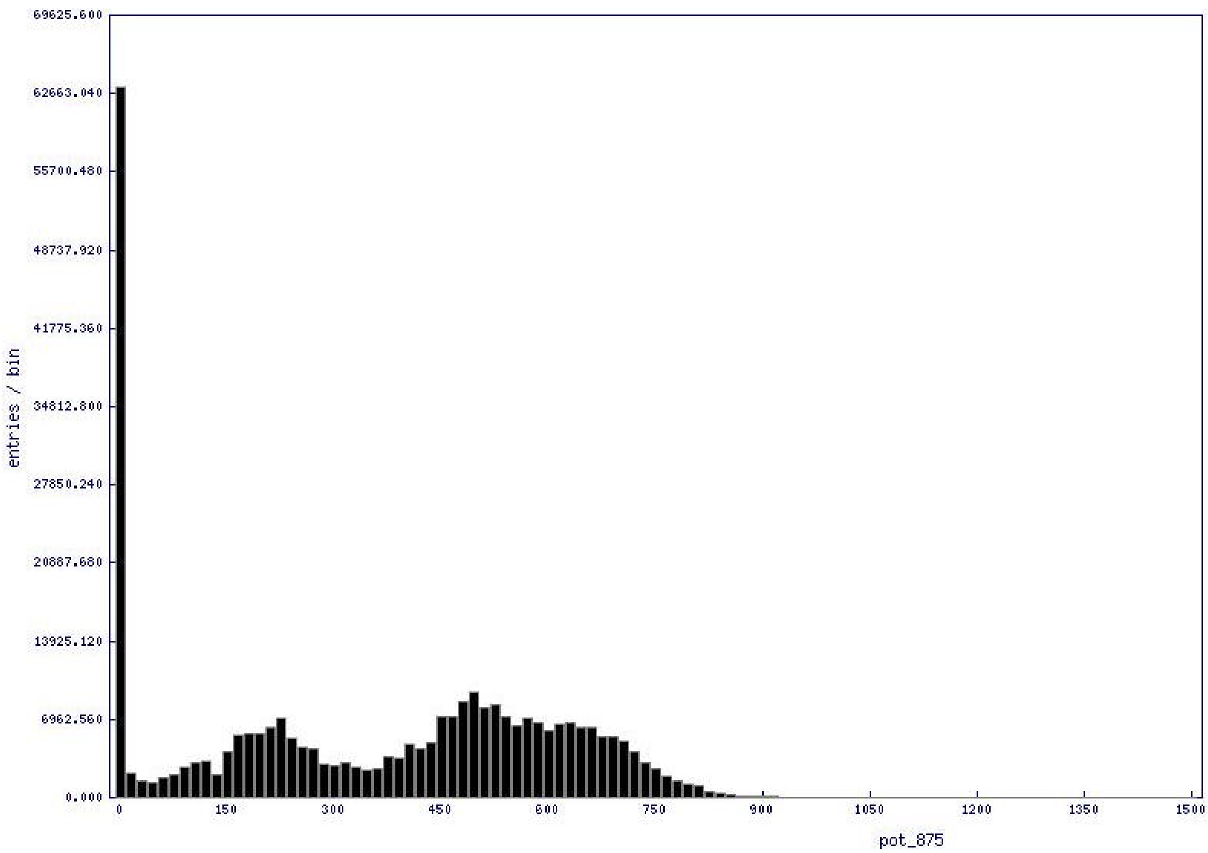
The technique described above works very well to match the distribution of minutes with 0,1,2 … neutrinos actually observed in the MiniBooNE experiment to date.

APPENDIX A
POT/minute for n=0,1,2 … 7.

This appendix contains a set of histograms of the number of minutes with n=0,1,2 … neutrinos binned by POT/minute. Each histogram is for a selected value of n.

The scale on all these plots is from zero to 1500 E12 POT using "pot_875" from boodb_anal_pot which is defined as "pot e12, toroid 875".

For n=0



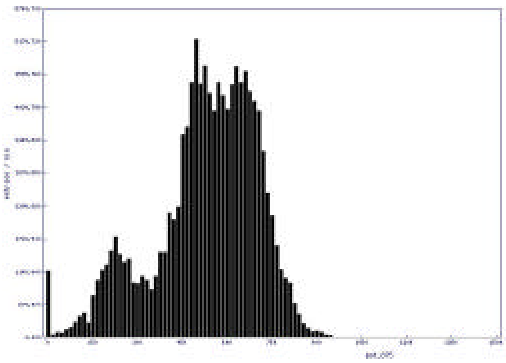
There are four "features" of this histogram worth noting.

1) There are a large number of entries at or very close to zero.
2) There is a peak near ~200E12.
3) There is another peak near ~500.
4) There is a fourth peak (probably) near ~650.

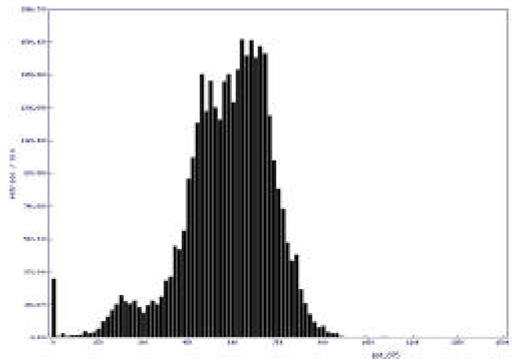
These features are present to a lesser degree in the following histograms, which show n not equal to zero. In particular note the trend for higher n to correlate with smaller peaks at the lower values of pot_875. This is not unexpected, but it is nice to see it in the data.

POT/minute for n=1,2 … 7.

n=1

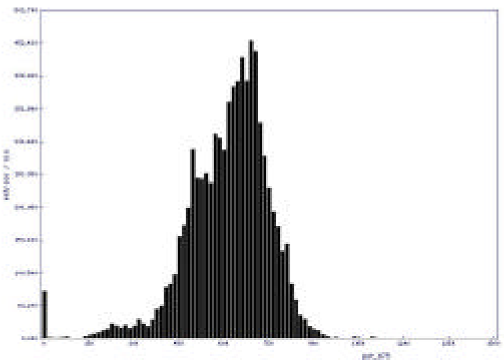


n=2

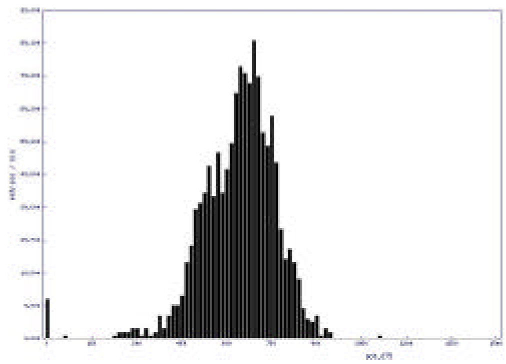


---

n=3

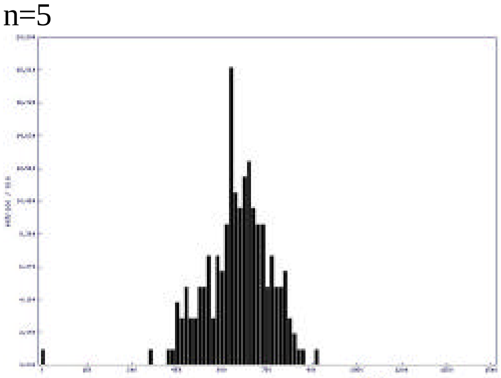


n=4

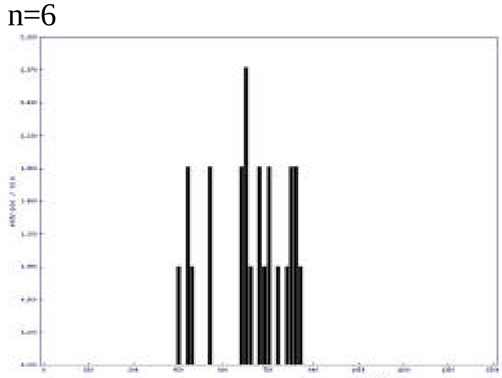


---

n=5

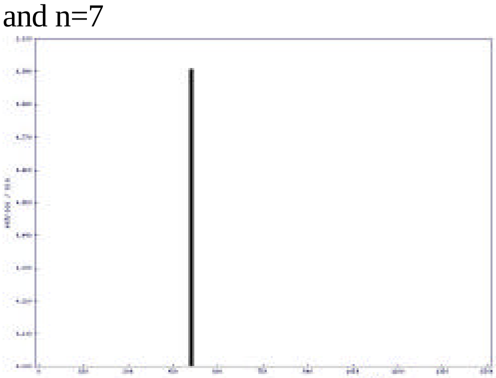


n=6



---

and n=7



Footnotes

---

[i] Thanks to Steve Brice for keeping the conceptual development on a productive path.

[ii] Thanks to Jocelyn Monroe for discussions clarifying some of the peculiarities of the MiniBooNE data and the experiment's setup, and for lending me her copy of Bevington.

[iii] Thanks to Gordon McGregor for filling this table and for teaching me most of everything I know (so far) about how to use the Analysis Tables.

[iv] See for example, "Data Reduction and Error Analysis for the Physical Sciences" (1969), by Philip R. Bevington.

[v] This is the same cut described in the MiniBooNE memo Neutrinos/POT for the MiniBooNE Week Ending on November 16, 2003, David Finley, 04/21/04.

[vi] Ibid.

[vii] Here the word "pulse" refers to a single Booster batch which arrives at the MiniBooNE target with ~80 bunches distributed over ~1.6 microseconds. These pulses occur at a 15 Hz rate in short bursts with the maximum average rate limited to 5 Hz by the MiniBooNE horn. At the moment the average rate if further limited by uncontrolled losses in the Booster.

[viii] http://www-boone.fnal.gov/publicpages/progress_monitor.html .